A Novel Approach for Automatic Image Annotation and Retrieval

T. Sumathi[1], Dr. M. Hemalatha[2]

Abstract :

The development of technology generates huge amounts of non-textual information, such as images. An efficient image annotation and retrieval system is highly desired. Clustering algorithms make it possible to represent visual features of images with finite symbols. Based on this, many statistical models, which analyze correspondence between visual features and words and discover hidden semantics, have been published. This application of computer vision technique is used in image retrieval system to organize and locate images of interest from a database. In this work, we introduce an innovative hybrid model for image annotation that treats annotation as a retrieval problem. The proposed technique utilizes low level image features and a simple combination of basic distances using JEC to find the nearest neighbors of a given image; the keywords are then assigned using SVM approach which aims to explore the combination of three different methods. First, the initial annotation of the data using two known methods, and that takes the hierarchy into consideration by classifying consecutively its instances; finally, we make use of pair wise majority voting between methods by simply summing strings in order to produce a final annotation. The proposed technique results show that this outperforms the current state of art methods on the standard datasets.

[1]Research Scholar, Karpagam University, Coimbatore.

[2]Asst. Prof & Head, Department of software systems, Karpagam University, Coimbatore

Keywords : *JEC, SVM, image annotation, image retrieval, radial basis function.*

I. Introduction

As high resolution digital cameras become more affordable and widespread the high quality digital image becomes ever more available and useful. With the exponential growth on high quality digital images, there is an urgent need to support more effective image retrieval over large scale archives. However content based image retrieval(CBIR) is still in its infancy and most existing CBIR systems can only support feature based image retrieval. Unfortunately, the naïve users may not be familiar with low level visual features and it is very hard for them to specify their query concepts by using low level visual features directly. Thus there is a great need to develop automatic image annotation framework, so that the naïve users can specify their query concepts easily by using the relevant keywords. However the performance o f image classifiers depends on two inter related issues: (1) suitable frameworks for image content representation and automatic feature extraction. (2). Effective algorithm for image classifier training and feature subset selection.

To address the first issue there are two widely accepted approaches for image content representation and feature extraction. To address the second issue for automatic image annotation two approaches are widely used to train the image classifiers. (a) Model based approach by using Gaussian mixture model to approximate the underlying distribution of image classes in the high dimensional

feature space (b) SVM-based approach by using support vector machine(SVM) to directly learn the maximum margins between the positive images and the negative images. In this work, SVM based approach is used to enable more effective classifier training with small generalization error rate in high dimensional feature space. However, searching the optimal methods (i.e. SVM parameters) is very expensive and its performance is very sensitive to the adequate choice of kernel function. So, for the annotation process we relied on SVM with a Radial basis function (RBF) kernel due to its performance. In this paper, we have proposed a hybrid hierarchical framework by incorporating the feature hierarchy and boosting to scale up SVM image classifier training. This framework is done in Mat lab using the popular label me web based annotation tool implementation.

II. Related Work

A large number of techniques have been proposed in the last decade [1]. Most of these treat annotation as translation from image instances to keywords. The translation paradigm is typically based on some model of image and text co-occurrences [1]. Latent Dirichlet Allocation (Corel LDA) [1] considers association through a latent topic space in a generatively learned model. Mori et al. [4] used a Co-occurrence Model in which they looked at the co-occurrence of words with image regions created using a regular grid. Monay and Gatica-Perez [4] introduced latent variables to link image features with words as a way to capture co-occurrence information. The addition of a sounder probabilistic model to LSA resulted in the development of probabilistic latent semantic analysis (PLSA) [4]. Blei and Jordan [4] viewed the problem of modeling annotated data as the problem of modeling data of different types where one type describes the other. Jeon et al. [4] improved on the result of Duygulu et al. by introducing a generative language model referred as Cross Media Relevance Model (CMRM) the same process used by Duygulu et al. was chosen to calculate the blob representation of images. They assumed that this could be viewed as analogous to the cross-lingual retrieval problem to perform both image annotation and ranked retrieval. Lavrenko et al. [4] argued that the process of quantization from continuous image features into discrete blobs, as the approach used by the machine translation model and the CMRM model, will cause the loss of useful information in image regions. While Feng et al. [4] modified the above model such that the probability of observing labels given an image was modeled as a multiple-Bernoulli distribution. In addition, they simply divided images into rectangular tiles instead of applying automatic segmentation algorithms. Their Multiple Bernoulli Relevance Model (MBRM) achieved further improvement on performance. Liu.et. al. [4], they estimated the joint probability by the expectation over words in a pre-defined Lexicon. It involves two kinds of critical relations in image annotation. First is the word-to-image relation and the second is the word-to-word relation. Torralba and Oliva [4] focused on modeling a global scene rather than image regions. This scene-oriented approach can be viewed as a generalization of the previous one where there is only one region or partition which coincides with the whole image. Yavlinsky et. al. [4] followed an approach using global features together with robust non-parametric density estimation and the technique of kernel smoothing. Jin et.al [4] proposes a new frame work for automated image annotation that estimated the probability for language model to be use for annotation an image.

III. Data set description

In this method we have utilized flicker dataset which contains 550 images in which 90% has been considered as training dataset and 10% as testing dataset.

IV. Methodology

Annotation of images in this work undergoes several stages: first, we extract information from the images and form a feature vector, hence we train several SVMs to create a model from the data for annotation accordingly to the mentioned approaches, flat and axis-wise, and position wise approaches herein tested. Finally we use majority voting, by summing strings, for a pair wise fusion between all three methods. We treat image annotation as a process of transferring keywords from nearest neighbors. The neighborhood structure is constructed using simple low-level image features resulting in a rudimentary model. A general flowchart of our procedure can be found in Fig. 1.

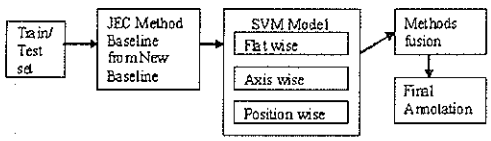

Figure.1. the frame work of our proposed system

4.1 Feature Extraction

To extract information from the images we used both global and a local image descriptor in a JEC approach. Feature selection was made accordingly to the desired image properties that we aimed to discriminate: color, texture and shape. All global descriptors were extracted using the Local and Web Image Retrieval Engine.

A. Color

RGB is the default color space for image capturing and display, both HSV and LAB isolate important appearance characteristics not captured by RGB. The RGB, HSV, and LAB features are 16-bin-per-channel histograms in their respective color spaces. To determine the corresponding L1 distance measures, as it performed the best for RGB and HSV, while KL-divergence was found suitable for LAB distances [1].

B. Combining distances

Joint Equal Contribution (JEC). If labeled training data is unavailable, or the labels are extremely noisy, the simplest way to combine distances from different descriptors would be to allow each individual distance to contribute equally (after scaling the individual distances appropriately). Let I_i be the i^{th} image, and say we have extracted N features $f_i^1, f_i^2, \ldots\ldots, f_i^N$. Let us define $d_{(i,j)}^k$ as the distance between f_i^k and f_j^k. We would like to combine the individual distances $d_{(i,j)}^k$, k = 1. . . N to provide a comprehensive distance between image I_i and I_j. Since, in JEC, each feature contributes equally towards the image distance, we first need to find the appropriate scaling terms for each feature. These scaling terms can be determined easily if the features are normalized in some way (e.g., features that have unit norm), but in practice this is not always the case. We can obtain estimates of the scaling terms by examining the lower and upper bounds on the feature distances computed on some training set. We scale the distances for each feature such that they are bounded by 0 and 1. If we denote the scaled distance as $d_{(i,j)}^k$ we can define the comprehensive image distance between images I_i and I_j as $\sum_{k=1}^{N} \frac{d_{(i,j)}^k}{N}$. We refer to this distance as Joint Equal Contribution (JEC).

4.2. Annotation

For the annotation process we relied on SVM's with a Radial Basis Function (RBF) kernel due to their

performance in the 2007-2009 Image CLEF medical image annotation tasks. In order to design speedy image retrieval systems, we use the SVM. The SVM [9] first maps the data into a higher dimensional input space by some kernel functions and then learns separating hyperspaces to maximize the margin. Currently, because of its good generalization capability, this technique has been applied in many areas. In our experiments, the RBF Kernel

$$K(x_1\text{-}x_2) = \exp(-|x1\text{-}x2|^2/\sigma^2)$$

is selected as the kernel function. So there is a corresponding parameter σ, to be tuned. A large value of ^{2} indicates a stronger smoothing. Moreover, there is another parameter γ, needing tuning to find the tradeoff between to stress minimizing of the complexity of the model and to stress good fitting of the training data points. We have set up a framework in MATLAB using the popular label me web based implementation. We performed an extensive grid-search on the common approaches to this problem, flat and axis-wise strategies, to optimize the kernel parameters using 10-fold cross validation. Each image is classified one axis at the time but, unlike the axis-wise method, conceptualization of the image content does not take the full meaning of the axis into consideration. Instead, we first consider the highest hierarchical position of the axis, its root, and use the whole training set to perform an initial classification. Afterwards, we reduce the training set to those images which match the initial classification, a semantic reduction of the training set, and classify the hierarchically subsequent inferior position. We undergo this top-down process thorough the axis tree until it is completely classified. We undertake the same methodology for all axes and assemble the final annotation. After the annotation from the three methods separately we make pair wise fusions of these by summing strings. The chart given below shows the percentage of keywords being annotated in our flicker dataset.

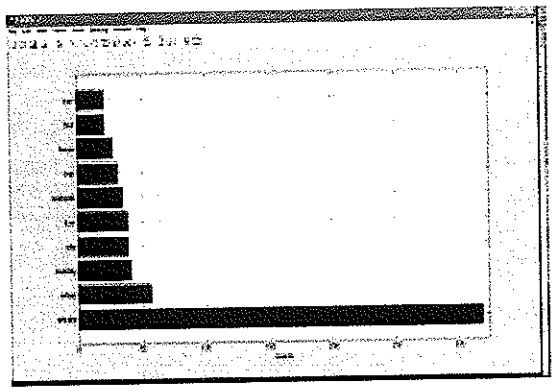

Figure: 2 Chart showing the annotation statistics

V. Evaluation and Discussion

5.1 Evaluation of annotation.

To evaluate annotation, we query images from the test dataset using 20 frequent keywords from the vocabulary. The image will be retrieved if the automatically established annotation contains the query keyword. We evaluate the result using P% and R% denotes the mean precision and the mean recall, respectively, over all keywords in percentage points. N+ denotes the number of recalled keywords. Note that the proposed simple baseline technique (JEC) [1] outperforms state-of-the-art techniques in all datasets [2]. The precision, recall and common E measure which are defined as

$P = NUM_{correct} / NUM_{retrieved}$

$R = NUM_{correct} / NUM_{exists}$, and

$$E(p,r) = 1 - 2/([1/p] + [1/r])$$

Table 1 : Evaluation measure of our proposed method.

Methods	P%	R %	N+
RGB	18	22	110
RGB16	12	14	94
HSV	17	19	80
HSV16	14	16	108
LAB	12	13	102

Table 2 : Comparison of other feature extraction methods with JEC

Methods	P%	R %	N+
JEC+SVM	19	22	110
Lasso+SVM	12	19	94
group lasso + SVM	10	18	87
Least Square + SVM	10	13	88
L2 regularization + SVM	11	14	93

Table 3 : Comparison with other annotation methods

Methods	Overall Precision	Recall	E measure
Baseline (greedy app)	0.20	0.23	0.786
Hierarchical model	0.34	0.29	0.636
Proposed method	0.771	0.356	0.513

5.2 Discussion

We have evaluated our method based on various feature selection like RGB, HSV and LAB. The table 1 shows the performance our method based on various features and from the results it can be concluded that JEC when combined with RGB feature performs well. The results of table 2 clearly show that JEC when combined with SVM outperforms all other methods like lasso, group lasso, least square and L2 regularization method. The results of table 3 show that our method has higher precision and recall rate compared with the other two methods. (i.e., New base line method using greedy approach and hierarchical method using bag of words approach).

Conclusion

Our SVM model on JEC works well, with good performances, needing much less training time than other systems. It could be concluded that our system with JEC feature is efficient for this task. The goal of our work was to develop a new annotation method by combining the JEC distance measure with that of the hierarchical method for image annotation. Experiments on these dataset reaffirm the enormous importance of considering multiple sources of evidence to bridge the gap between the pixel representations of images and the semantic meanings. It is clear that a simple combination of basic distance measures defined over in extraction of image features has effectively served us to provide a better annotation results.

References

[1] A New Baseline for Image Annotation Ameesh Makadia, Vladimir Pavlovic and Sanjiv Kumar, In Computer Vision – ECCV 2008, Vol. 5304 (2008), pp. 316-329.

[2] Igor F. Amaral, Filipe Coelho, Joaquim F. Pinto da Costa and Jaime S. *'Cardoso Hierarchical Medical Image Annotation Using SVM-based Approaches'* 978-1-4244-6561-3/10/$26.00 ©2010 IEEE

[3] T. Sumathi, M. Hemalatha *'An Empirical Study on Performance Evaluation in Automatic Image Annotation and Retrieval'* published in International journal of advanced research in computer science, Vol.1., No.4.,Nov-dec-2010

[4] T. Sumathi, C. Lakshmi Devasena, and M. Hemalatha *'An Overview of Automated Image Annotation Approaches'* International Journal of

Research and Reviews in Information Sciences Vol. 1, No. 1, March 2011

[5] Nasullah Khalid Alham, Maozhen Li1, Suhel Hammoud and Hao Qi, *'Evaluating Machine Learning Techniques for Automatic Image annotations'* ieeexplore.ieee.org/iel5/pp53-58

[6] Syaifulnizam Abd Manal, MD Jan Nordin *'Review on statistical approaches for automatic image annotation, International conference on electrical engineering and informatics'*, 978-1-4244-4913-2/2009 IEEE

[7] Nasullah Khalid Alham, Maozhen Li, Suhel, *'evaluating Machine learning techniques for automatic image annotations'* 978-0-7695-3735-1/09 2009 IEEE

[8] YuliGao et.al, *'Automatic image annotation by incorporating feature hierarchy and Boosting to scale up SVM classifiers'*, ACM Multimedia, October 22-28,2006

[9] Herve Glotin, Zhong-Qiu Zhao, Emilie Dumont, Thes is for rehabilitation of research direction, university Sod Toulon-Var, Toulon(2007)

[10] Herve Glotin, H., Zhao, Z.Q., Ayache, S., *'Efficient image concept indexing by harmonic and arithmetic profiles entropy'*, 2009 IEEE international conference on image processing, Nov 7-11, 2009.

Author's Biography

Dr. M. Hemalatha completed MCA MPhil., PhD in Computer Science and Currently working as a Asst Professor and Head, dept of software systems in Karpagam University. Ten years of Experience in teaching and published Twenty seven paper in International Journals and also presented seventy papers in various National conferences and one international conferences Area of research is Data mining, Software Engineering,bioinformatics, Neural Network. Also reviewer in several National and International journals.

T. Sumathi is presently doing Ph.D in Karpagam University, Coimbatore, Tamilnadu, India and has completed M.Phil (computer science) in the year 2006 and MCA degree in 1999 and B. Sc(computer science) in 1996. Major research area is Image processing and title for the research is image annotation. At Present working as Lecturer in karpagam University, Coimbatore.